

## REMARKS

The Final Office Action dated May 6, 2009 has been received and reviewed. Claims 1-15 and 17-28 are pending in the subject application. Each of claims 1-4, 7, 14, 19, 20 and 22-28 has been amended herein. Care has been exercised to introduce no new matter. Applicants respectfully request reconsideration of the subject application in view of the above amendments and the following remarks.

### Rejections based on 35 U.S.C. § 103(a)

#### A) Applicable Authority

Title 35 U.S.C. § 103(a) declares that a patent shall not issue when “the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains.” In *Graham v. John Deere*, the Supreme Court counseled that an obviousness determination is made by identifying: the scope and content of the prior art; the level of ordinary skill in the prior art; the differences between the claimed invention and prior art references; and secondary considerations. *See Graham v. John Deere Co.*, 383 U.S. 1 (1966).

“In determining the differences between the prior art and the claims, the question under 35 U.S.C. 103 is not whether the differences themselves would have been obvious, but whether the claimed invention as a whole would have been obvious.” MPEP § 2141.02(I) (emphasis in original) (citing *StratoFlex, Inc. v. Aeroquip Corp.*, 713 F.2d 1530, 218 USPQ 871 (Fed. Cir. 1983)).

“The examiner bears the initial burden of factually supporting a *prima facie* conclusion of obviousness. If the examiner does not produce a *prima facie* case, the applicant is

under no obligation to submit evidence of nonobviousness . . . . To reach a proper determination of obviousness, the examiner must step backward in time and into the shoes worn by the hypothetical ‘person of ordinary skill in the art’ when the invention was unknown and just before it was made. In view of all factual information, the examiner must then determine whether the claimed invention ‘as a whole’ would have been obvious at that time to that person. *Id* (emphasis added). Knowledge of applicant’s disclosure must be put aside in reaching this determination . . . . [I]mpermissible hindsight must be avoided and the legal conclusion must be reached on the basis of the facts gleaned from the prior art.” MPEP § 2142.

“The key to supporting any rejection under 35 U.S.C. 103 is the **clear articulation of the reason(s)** why the claimed invention would have been obvious.” MPEP § 2142 citing *KSR Int’l Co. v. Teleflex Inc.*, 127 S. Ct. 1727 (U.S. 2007) (emphasis added), which notes that the analysis supporting a rejection under 35 U.S.C. 103 should be made explicit.

B.) Rejections Based Upon Obata in view of Najork

Claims 1-15, 17-18, 22-23 and 26-28 have been rejected under 35 U.S.C. § 103(a) as being unpatentable over EP 1120717 to Obata et al., (hereinafter “Obata”) in view of U.S. Patent No. 6,263,364 to Najork et al. (hereinafter “Najork”). As the asserted combination of references fails to teach or suggest all of the limitations of the rejected claims, as amended herein, Applicants respectfully traverse the rejection, as hereinafter set forth.

With regard to **claim 1**, it is asserted in the Office Action that Obata teaches “an indexer that places items into a chunk, wherein the items are the results returned by a web crawl (‘The gatherer process 304 inserts the starting URLs 306 into a transaction log 310. The transaction log 310 identifies those documents that are to be crawled during the current crawl’- See [0036]).”<sup>1</sup>

The transaction log taught by Obata does not correspond to a “chunk” as claimed by Applicants. Obata’s transaction log identifies documents that are *to be* crawled. “The transaction log 310 identifies those documents that are *to be* crawled during the current crawl.”<sup>2</sup> “A current document at each document address specification listed in the transaction log is retrieved from its Web site and processed.”<sup>3</sup> The transaction log does not contain results of a web crawl. In contrast, independent claim 1, as amended herein, recites “an indexer that places items with similar properties into respective chunks,...wherein the items are *the results returned by a web crawl*.” Thus, Applicants’ “chunk” contains items that are the *results* of a webcrawl.

Further, **claim 1** has been amended to recite “wherein the respective chunks include at least one rank chunk and at least one webmap chunk.” The transaction log taught by Obata is not a “rank chunk” nor a “webmap chunk” as claimed by Applicants, nor have Applicants been able to find any portion of Obata that teaches or suggests “wherein the respective chunks include at least one rank chunk and at least one webmap chunk.”

With regard to **claims 1, 22 and 26**, it is asserted in the Office Action that Obata teaches “the properties are at least one of average time between change and average importance of documents in the respective chunk (‘Historical information such as the first access time 422,

---

<sup>1</sup> *Office Action of 5/6/2009*, p. 2.

<sup>2</sup> *Obata, ¶0036.*

<sup>3</sup> *Obata, ¶0021.*

the last access time 424, the change count 426, and the access count 428 are used in a statistical model for deciding if a document should be accessed during an adaptive incremental crawl' - See [0046]; 'In accordance with other aspects of the invention, each Web crawl begins with an active probability distribution containing a plurality of probabilities indicative that a document has changed at a given change rate' See [0010])."

**Claims 1, 22 and 26** have been amended to recite "wherein the properties *include* average time between change and average importance of documents." Applicants submit that Obata teaches the number of intervals between changes, but not an average time between change. "The average time between accesses (intervals) is computed and stored as the interval time (DT). The *number of intervals* between changes is calculated (NC)."<sup>4</sup> The average taught by Obata is the average time *between accesses*, not between changes. Obata assumes that changes occur at equal intervals: "assumption that the changes recorded in the change count 426 occur at equal intervals during the accesses 1618."<sup>5</sup> Thus, Obata does not teach or suggest "wherein the properties include average time between change" as claimed by Applicants. Nor have Applicants found any portion of Obata that teaches or suggests "wherein the properties include...average importance of documents" as claimed by Applicants.

C. Rejections Based Upon Obata in view of Najork and Further in view of Dean

**Claims 5 and 8** have been rejected under 35 U.S.C. § 103(a) as being unpatentable over Obata in view of Najork and further in view of U.S. Patent No. 7,305,610 to Dean et al. As the asserted combination of references fails to teach or suggest all of the

---

<sup>4</sup> *Obata*, ¶0101.  
<sup>5</sup> *Obata*, ¶0101.

limitations of rejected claims 5 and 8, Applicants respectfully traverse the rejection, as hereinafter set forth.

Regarding **claim 5**, it is asserted in the Office Action that Dean teaches “facilitating load balancing among a plurality of crawlers (‘FIG. 7 shows a flow chart of a process of adjusting stall times’- See Col. 7, lines 34-35; ‘Once the actual retrieval time is determined, the stall time for the selected host can be adjusted according to the retrieval time at a step 503’- See Col. 7, lines 40-43; ‘each computer system I can be executing one or more web crawler that traverses hyperlinked documents and saves information regarding the traversed hyperlinked documents on the computer system’- See Col. 3, lines 60-63).”<sup>6</sup>

Applicants respectfully submit that Dean does *not* teach facilitating “load balancing amongst a plurality of crawlers” as claimed by Applicants. The “stall times” taught by Dean are a means of providing rate limiting *of hosts* that prevents a host from being crawled too frequently. “[L]ink server 201 attempts to ensure that each particular host is *not crawled too frequently*.”<sup>7</sup> “In order to accomplish *rate limiting of hosts*, each *host* has an associated *stall time*, which is the earliest time at which another link from this host should be crawled.”<sup>8</sup> Stall times prevent excessive crawling of a host, but do not “facilitate load balancing amongst a plurality of crawlers” as claimed by Applicants. In other words, as taught by Obata, a crawler is prevented from crawling a host too frequently, but there is no balancing of loads between the crawlers.

Further, Applicants claim “a master control process that can *modify the chunk map to facilitate load balancing* amongst a plurality of crawlers.” The Office has provided no

---

<sup>6</sup> *Office Action of 5/6/2009*, p. 13.

<sup>7</sup> *Dean*, col. 4, ll. 6-18.

<sup>8</sup> *Dean*, col. 6, ll. 46-48.

argument that the art of record teaches that load balancing among crawlers can be facilitated by modifying the chunk map, nor “a *master control process that can modify the chunk map to facilitate load balancing*” as claimed by Applicants.

D.) Rejections Based Upon Dean in view of Evans

**Claims 19-21** have been rejected under 35 U.S.C. § 103(a) as being unpatentable over Dean in view of U.S. Publication No. 2004/0030683 to Evans et al. (hereinafter “Evans”). As the asserted combination of references fails to teach or suggest all of the limitations of rejected claims 19-21, Applicants respectfully traverse the rejection, as hereinafter set forth.

Regarding **claim 19**, the claim has been amended. Applicants have been unable to find any portions of Dean or Evans, either singly or in combination, that teach or suggest “selecting a first chunk from a group that includes index chunks, rank chunks, content chunks, recrawl chunks and webmap chunks,” as claimed by Applicants, as amended.

Further, Applicants have been unable to find any portions of Dean or Evans, either singly or in combination, that teach or suggest “the stored properties include average time between change and average importance and are shared by all items in the first chunk,” as claimed by Applicants, as amended.

Further regarding **claim 19**, it is asserted in the Office Action that Dean teaches “a chunk map that stores properties associated with the respective chunk stored in a chunk table is employed to determine the first chunk (‘In order to accomplish rate limiting of hosts, each host has an associated stall time, which is the earliest time at which another link from this host should be crawled or released to a crawler’ - See Col. 6, lines 47-50; The stall times are a property associated with a respective host. The stall time is used to determine which host, and thus, which

group of links (chunk) should be crawled next).<sup>9</sup> However, as discussed above with regard to claim 5, the “stall times” taught by Dean are employed to limit the rate at which a host is crawled in order to avoid placing an undue burden on the host. “For example, if the next 100 highest priority uncrawled links are all from the same host, the technique of always crawling the link with the highest priority will likely have the effect of placing *an undue strain on this host*....FIG.4 shows a block diagram of a distributed system that crawls hyperlinked documents and can *provide rate limiting* for hosts....link server 201 attempts to ensure that each particular host is *not crawled too frequently*.<sup>10</sup> The “stall time” taught by Dean is a means of providing rate limiting that prevents a website from being crawled too frequently. “In order to accomplish *rate limiting of hosts*, each *host* has an associated *stall time*, which is the earliest time at which another link from this host should be crawled.”<sup>11</sup> Thus, the stall time limits the rate at which a host should be crawled, but provides *no* information about properties of the links that reside on the host.

Further, Applicants have been unable to find any portions Dean or Evans, either singly or in combination, that teach or suggest “the stored properties include average time between change and *average importance* and are shared by *all items in the first chunk*,” as claimed by Applicants, as amended.

E.) Rejections Based Upon Obata in view of Najork and Dingsor

Claims 24-25 were rejected under 35 U.S.C. § 103(a) as being unpatentable over Obata et al. in view of Najork et al. and U.S. Patent No. 7,058,727 to Dingsor et al. (hereinafter “Dingsor”). As the asserted combination of references fails to teach or suggest all of the

---

<sup>9</sup> Office Action of 5/6/2009, pp. 14-15.

<sup>10</sup> Dean, col. 4, ll. 6-18.

<sup>11</sup> Dean, col. 6, ll. 46-48.

limitations of rejected claims 24 and 25, Applicants respectfully traverse the rejection, as hereinafter set forth.

Regarding **claim 24**, it is asserted in the Office Action that Obata teaches “the chunk comprising document files associated with one or more uniform resource locators.”<sup>12</sup> However, Applicants have been unable to find any portions of Obata, Najork or Dingsor, either singly or in combination, that teach or suggest “document files associated with one or more uniform resource locators, wherein properties of all the document files *include time between change and importance of a document*” as claimed by Applicants, as amended, for at least the reasons given above with regard to claim 1.

Further, Applicants have been unable to find any portions of Obata, Najork or Dingsor, either singly or in combination, that teach or suggest “wherein data in the data structure is compressed” as claimed by Applicants, as amended.

Claims 1, 19, 22, 24 and 26 are believed to be in condition for allowance for at least the reasons given above. Claims 2-15, 17-18, 20-21, 23, 25, and 27-28 depend either directly or indirectly from claims 1, 19, 22, 24 and 26, thus incorporating by reference all of the features of those claims, and are therefore also believed to be in condition for allowance for at least the reasons given above.

---

<sup>12</sup> *Office Action of 5/6/2009*, p. 17.

**CONCLUSION**

For at least the reasons stated above, claims 1-15 and 17-28 are now believed to be in condition for allowance. Applicants respectfully request withdrawal of the pending rejections and allowance of the claims. If any issues remain that would prevent issuance of this application, the Examiner is urged to contact the undersigned – 816-474-6550 or [twilhelm@shb.com](mailto:twilhelm@shb.com) (such communication via email is herein expressly granted) – to resolve the same.

The fee for a one-month extension of time and Request for Continued Examination are submitted herewith. It is believed that no additional fee is due. However, if this belief is in error, the Commissioner is hereby authorized to charge any amount required, or credit any overpayment, to Deposit Account No. 19-2112, referencing attorney docket number 306413.01/MFCP.149238.

Respectfully submitted,

/Tawni L. Wilhelm/

Tawni L. Wilhelm  
Reg. No. 47,456

TLW/MMS/kmp  
SHOOK, HARDY & BACON L.L.P.  
2555 Grand Blvd.  
Kansas City, MO 64108-2613  
816-474-6550